



INDICADORES DE CALIDAD DE DATOS ABIERTOS: EL CASO DEL PORTAL DE DATOS ABIERTOS DE BARCELONA

Open data quality metrics: Barcelona open data portal case



Alberto Abella, Marta Ortiz-de-Urbina-Criado y Carmen De-Pablos-Heredero



Alberto Abella es ingeniero superior de telecomunicaciones y doctor en Organización de Empresas. Ha publicado artículos en revistas de impacto como *El profesional de la información* y *Cities*. Es experto en los campos de *open data* y *open government* aplicados al sector público y privado. Trabaja en el desarrollo de políticas y herramientas de *open data* en *desideDatum Data Company*.
<http://orcid.org/0000-0003-0243-4606>

alberto.abella@meloda.org



Marta Ortiz-de-Urbina-Criado es doctora en Dirección y Organización de Empresas y profesora titular de universidad en la *Universidad Rey Juan Carlos*. Ha publicado artículos en revistas de impacto como *International business review*, *International journal of technology management*, *Journal of knowledge management*, *El profesional de la información* y *Cities*, entre otras.
<http://orcid.org/0000-0001-7527-6798>

marta.ortizdeurbina@urjc.es



Carmen De-Pablos-Heredero es doctora en Ciencias Económicas y Empresariales, profesora titular de universidad y directora del *Master Universitario en Organización de Empresas* en la *Universidad Rey Juan Carlos* y el *Master en Gestión de Proyectos Logísticos SAP*. Ha publicado artículos en revistas de impacto como *CEDE*, *Interciencia*, *TIBE*, *Experimental & clinical cardiology*, *Revista de economía mundial*, *El profesional de la información*, *International journal of environmental research & public health*, *Cities*, *Studies in higher education*, etc.
<http://orcid.org/0000-0003-0457-3730>

carmen.depablos@urjc.es

*Universidad Rey Juan Carlos, Facultad de Ciencias Jurídicas y Sociales
Paseo de los Artilleros, s/n. 28032 Madrid, España*

Resumen

Los datos abiertos son una infraestructura básica para la creación de negocios y de productos y servicios. Para hacer un análisis de su utilidad hay que tener en cuenta que no es igual el acceso a los mismos y su difusión que su reutilización. El objetivo de este trabajo es analizar como el modelo de cinco estrellas de Berners-Lee y otros factores ayudan a evaluar la calidad de los datos de cara a su reutilización en el portal de datos abiertos de Barcelona y analizar su relación con su descarga y sus temáticas. Los resultados obtenidos muestran que puede ser interesante incorporar aspectos como la frecuencia de actualización de datos y la geolocalización en los modelos que miden la calidad de los datos abiertos para su reutilización.

Palabras clave

Datos abiertos; Calidad; Reutilización de datos; Publicación de datos; Difusión de datos; Indicadores; Indicadores de calidad; Barcelona.

Abstract

Open data are a basic infrastructure for the creation of business, products, and services. To analyze their usefulness, the differences between data access, and dissemination, and reuse must be taken into account. The objective of this research is to analyze how Berners-Lee's 5-star model and other factors help evaluate the quality of data and its use in the Barcelona open data portal, including its relationship to downloads and themes. Results show that it may be of interest to include aspects such as data updating frequency and geolocation in the models that measure the quality of data for reuse.

Artículo recibido el 30-11-2017
Aceptación definitiva: 07-03-2018

Keywords

Open data; Quality; Data reuse; Data publication; Data dissemination; Indicators; Quality metrics; Barcelona.

Abella, Alberto; Ortiz-de-Urbina-Criado, Marta; De-Pablos-Heredero, Carmen (2018). "Indicadores de calidad de datos abiertos: el caso del portal de datos abiertos de Barcelona". *El profesional de la información*, v. 27, n. 2, pp. 375-382.

<https://doi.org/10.3145/epi.2018.mar.16>

1. Introducción

Los datos abiertos (*open data*) se pueden definir como

"aquellos datos que se publican en sitios web de acceso público con el fin de ser reutilizados y redistribuidos por aquellas personas u organizaciones que lo deseen, sin tener ningún tipo de restricciones para ello" (Oviedo; Mazón; Zubcoff, 2013).

Estos autores destacan que los portales de datos abiertos son los que ponen a disposición estos datos. Según Marcos-Martín y Soriano-Maldonado (2011) y Weiskopf y Weng (2013), los datos abiertos permiten que cualquier persona u organización que los utilice genere nueva información y servicios. Un ejemplo es *DBpedia*, una base de conocimiento abierta y gratuita que ofrece un servicio complementario para *Wikipedia* y que publica los datos utilizando estándares abiertos como URIs, http, html, RDF, *Sparql*, y licencias de datos abiertos (Auer et al., 2007).

<http://wiki.dbpedia.org/about>

Sin embargo una cosa es la cantidad de datos abiertos disponibles y otra su calidad. El éxito de los sistemas de datos abiertos depende, más que del simple acceso, de la valoración de la calidad de los mismos (Janssen; Charalabidis; Zuiderwijk, 2012). Ferrer-Sapena, Peset y Alexandre-Benavent puntualizan que

"la puesta a disposición de los datos no tiene sentido si posteriormente no se han desarrollado acciones de reutilización" (Ferrer-Sapena; Peset; Alexandre-Benavent, 2011, p. 263).

También comentan que hay que diferenciar entre el acceso y las posibilidades de su reutilización.

“ Los datos abiertos ofrecen muchas oportunidades para la creación de negocios ”

La mera existencia de datos abiertos no implica su reutilización, tal y como muestran las cifras de uso de algunos portales de datos abiertos. Por ejemplo, el portal paneuropeo de datos abiertos que se creó en noviembre de 2015 y que agregaba más de medio millón de metadatos de conjuntos de datos, tuvo en 2016 menos de 11 visitas al día. Por el contrario, el portal de la *Empresa Municipal de Transportes (EMT)* de Madrid ofrece información a tiempo real y tiene picos de más de 5 millones de conexiones diarias a su API para recoger información.

<http://europeandataportal.eu>

<http://opendata.emtmadrid.es>

Maseda-Seco, Bueno-de-la-Fuente y Méndez (2017) apuntan que

"no toda la información pública disponible o publicada en la Web es información abierta válida para su reutilización".

Además de publicar los datos, hay que dar acceso a los mismos y dotarlos de una estructura que permita su reutilización (Garriga-Portolà, 2011). Gértrudix, Gertrudis-Casado y Álvarez-García (2016) explican que hay cierta desconfianza de los ciudadanos respecto de las acciones que realizan las instituciones públicas para acceder, utilizar y reutilizar datos abiertos. Maseda-Seco, Bueno-de-la-Fuente y Méndez (2017) consideran que la clasificación de Berners-Lee (2006) es una herramienta útil para analizar la apertura y potencial de reutilización, explotación y consumo de los datos.

En el presente trabajo se aplica el modelo de las cinco estrellas de Berners-Lee para evaluar la calidad de los datos abiertos del portal de datos abiertos de la ciudad de Barcelona y analizar su relación con las descargas de datos y sus temáticas. El objetivo es encontrar vías que permitan evolucionar este modelo incorporando aspectos relevantes que no se han contemplado como frecuencia de actualización y geolocalización de los datos. Con este estudio se ofrece información útil para mejorar los procesos de publicación de datos abiertos, alinearlos con la demanda y facilitar su reutilización.

2. Calidad de los datos abiertos: las cinco estrellas de Berners-Lee

Sowe y Zettsu (2015) presentan un modelo de desarrollo de los datos abiertos con cinco fases:

- revisión, acceso, uso y reutilización;
- creación de nuevos conjuntos de datos;
- evaluación y selección;
- transformación;
- registro y publicación.

En este trabajo nos centramos en la primera fase y en los indicadores de medida para su acceso, uso y reutilización.

Muirragui-Irrazábal et al. (2016) resumen algunos de los indicadores de uso de las plataformas de datos abiertos:

- acceso: número de visitantes, entradas, salidas, páginas vistas, tiempos, tasas de abandono, comentarios y contenidos compartidos...;
- utilización: número de descargas, descargas por tipo de distribución, acceso y uso de los datos enlazados...;
- gestión: número de conjuntos de datos actualizados, distribución de los conjuntos publicados, frecuencia media de actualización de los datos, formatos utilizados, relación entre solicitudes de información y datos publicados...

La difusión y uso de datos implican la transferencia de la información de los sistemas internos a las herramientas de acceso a los datos abiertos, desde la búsqueda de datos hasta la descripción de características como autoría, acceso, frecuencia, tamaño, descripción del modelo de negocio, antes de su publicación (Janssen; Charalabidis; Zuiderwijk,

Tabla 1. Niveles de calidad para datos abiertos

Estrellas	Significado y ejemplos	Ventajas (V) e inconvenientes (I)
★	Los datos están disponibles en la Red independientemente del formato utilizado. Ej.: se publica como una tabla dentro del propio portal	V: datos accesibles en un documento desde cualquier licencia abierta. I: Dificultad para extraer datos del documento.
★★	Los datos se publican en la Web en un formato estructurado. Ej.: los datos son exportables en formato <i>Excel</i> .	V: los datos son accesibles de forma estructurada. I: se requiere un software propietario para extraer datos.
★★★	Los datos están publicados bajo un formato estructurado no propietario. Ej.: los datos son exportables en formato CSV.	V: los datos están disponibles en la web y se pueden utilizar. I: los datos no están integrados en la web.
★★★★	Los datos se identifican mediante URLs de manera que sean fácilmente interpretables. Ej.: los datos son exportables en formato html y otros.	V: los datos tienen una URI y pueden compartirse en la web. I: no están enlazados con otros datos.
★★★★★	Los datos están vinculados con otros datos de manera que se encuentran contextualizados. Ej.: los datos son exportables en una variedad de formatos y accedidos a través de hipervínculos a diferentes localizaciones.	V: el efecto red es posible porque están integrados y relacionados con otros datos en la Web. I: sólo se relacionan con datos con los que se ha preestablecido hipervínculo.

Fuente: adaptado de **Oviedo, Mazón y Zubcoff (2013)**

2012), e incluso, para los conjuntos de datos que se obtienen de información semántica (*linked data*), puede incluir enriquecimiento y transformación de datos (Zuiderwijk; Jeffery; Janssen, 2012).

Por otra parte, **Abella, Ortiz-de-Urbina-Criado y De Pablo-Heredero (2014; 2017a)** han elaborado la métrica *Meloda* que permite evaluar el grado de reutilización de los datos abiertos. Esta métrica se ha utilizado por otros autores como **Vicente-Paños y Jordán-Alfonso (2017)** para estudiar el acceso a la información del sector público y su reutilización en las comunidades autónomas de España. Estos autores observaron que cuando se calificaban a través de *Meloda* los portales de datos abiertos de las comunidades autónomas españolas, las puntuaciones más bajas se obtenían en las dimensiones de información geolocalizada e información en tiempo real.

En esta fase además del acceso, uso y reutilización hay que tener en cuenta la calidad de los datos abiertos. Uno de los modelos más conocidos de valoración de la calidad de los datos abiertos es el propuesto en mayo de 2010 por **Berners-Lee (Kalampokis; Tambouris; Tarabanis, 2011)**.

El modelo de las cinco estrellas de **Berners-Lee** es fruto de la necesidad evidenciada de forma temprana de ordenar datos que se encuentran deslocalizados en la Web.

Brin y Page (1998) resaltan que uno de los problemas que presentan las aplicaciones que permiten ordenar datos es que no son precisas y consumen demasiado tiempo.

Bizer, Heath y Berners-Lee (2009) describen diferentes prácticas que los proveedores de datos realizan para acceder a los datos. Ponen de manifiesto cómo la información semántica (*linked data*) tiene el potencial de revolucionar la forma en que se acceden y utilizan los datos. Los desarrolladores

se enfrentan al desafío de escalar la forma en la que se accede a modelos de datos dispersos y heterogéneos.

Martínez-Méndez, López-Carreño y Pastor-Sánchez (2014) realizan un análisis de los datos abiertos en el portal *datos.gob.es* y reconocen que la distribución de conjuntos de datos abiertos es muy diversa y que precisa de agrupación en meta-categorías para su estudio y análisis.

Hogan y Shepherd (2015) describen que dependiendo de a quién pertenezcan los datos, se han regulado de forma asimétrica, y ofrecen recomendaciones para conseguir un acceso más democrático a los datos.

Tambassi (2016) analiza la importancia de *dotar de sentido común* a las estructuras de datos, teniendo en cuenta su topología y su localización espacial y realiza un análisis de las principales ontologías.

El modelo fija cinco niveles de calidad para datos abiertos (cinco estrellas) poniendo el énfasis en dos aspectos relevantes (tabla 1 y figura 1):

- disponibilidad (niveles 1, 2 y 3);
- capacidad de reutilización (niveles 4 y 5).



Figura 1. Esquema del modelo de las cinco estrellas
<http://5stardata.info>

Yang, Lin y Yu (2015) utilizan este modelo para verificar la coherencia de diferentes conjuntos de datos dentro de la misma categoría en portales en Taiwan, mostrando su utilidad para este propósito.

Čerba y Jedlička (2016) buscan como experimento relaciones del concepto “bosque” y encuentran que el modelo se podría mejorar para establecer comparaciones con conceptos geográficos.

Gerhard, Schwerdel y Müller (2017) lo aplican para presentar un protocolo novedoso de transporte que se optimiza por sistemas de detección definidos en redes de software y se pueden redirigir datos a través de procesadores en red.

Piedra et al. (2017) analizan cómo este modelo presenta un marco adecuado para buscar, descargar y reutilizar datos de estudiantes y profesores, lo que permite aumentar los contenidos educativos y mejorar relaciones entre alumnos y profesores.

El modelo de las cinco estrellas de Berners-Lee es fruto de la necesidad de ordenar datos que se encuentran deslocalizados en la Web

Martínez-Méndez; López-Carreño y Pastor-Sánchez (2014) hacen un análisis (por sectores, tipo y naturaleza de los datos) del catálogo de datos abiertos del sector público español disponible en el portal *datos.gob.es*. También consideran para cada sector las aplicaciones que se han desarrollado a partir de estos datos y se identifican las diferencias entre sectores.

A pesar de la aceptación del modelo y su uso en diferentes sectores y áreas, este modelo presenta algunas limitaciones. **Oviedo, Mazón y Zubcoff (2013)** proponen el enriquecimiento del modelo de cinco estrellas con seis medidas: disponibilidad, capacidad de reutilización, relevancia, reputación, granularidad y visualización, y lo aplican al caso de la municipalidad de Pérez Zeledón en Costa Rica.

Shankar (2015) hace hincapié en incluir algún tipo de atributo que permita seleccionar datos de calidad ante la abundancia de datos. Propone considerar la utilización de narrativas e historias que permitan valorar, abrir y avalar la confianza de los datos.

Ribeiro y Vivacqua (2013) describen las dificultades de la reutilización de datos y la necesidad de incorporar métodos que permitan seleccionarlos en base a varios criterios. Para ello proponen la utilización de módulos de inferencia reutilizables, creados a partir de estándares de la web semántica, que hacen más sencillo utilizar inferencias y cálculos basados en datos.

Otras métricas para analizar el grado de reutilización de datos como *Meloda* (**Abella; Ortiz-de-Urbina-Criado; De-Pablos-Heredero, 2017a**) incorporan dimensiones como la geolocalización y la actualización de los datos que pueden ser consideradas para mejorar la evaluación de la calidad de los datos abiertos del modelo de las cinco estrellas.

3. Metodología

Teniendo en cuenta los estudios previos que han intentado enriquecer el modelo de las cinco estrellas, en este trabajo se lleva a cabo su aplicación en un escenario real para ofrecer orientaciones que permitan mejorarlo. Se ha realizado un análisis de los datos abiertos disponibles en el portal de datos abiertos de una de las principales ciudades inteligentes de España y Europa: Barcelona. A lo largo del mes de octubre de 2017 se realizó la recogida y el análisis de los datos.

El portal tenía el 10 de octubre de 2017 un total de 425 conjuntos de datos con 3.002 distribuciones. Sin embargo parte de ellos están alojados en webs y recursos externos, por lo que es necesario filtrarlos y así se reducen a 358 distribuciones provenientes de 192 conjuntos de datos. Sobre éstos se obtuvo información de los siguientes aspectos:

- Para evaluar las estrellas para cada conjunto de datos, se utilizó el campo *Openness_score* disponible en el portal mediante el acceso vía API. En este campo se dispone de la valoración de las estrellas para cada distribución de cada conjunto de datos.
- Para evaluar las descargas se utilizó el campo Descargas, extraído desde el API para cada distribución de cada conjunto de datos. Estas descargas sólo incluyen las directas, pero no las realizadas vía API.
- Para evaluar la temática de cada conjunto de datos, se utilizó el campo *Temática_NTI* disponible para cada conjunto de datos conforme a clasificación de temáticas del anexo IV de la norma técnica de interoperabilidad de reutilización de recursos de la información (NTI-RISP) (*España, 2013*).

La semantización es una tarea pendiente y para su implementación se requieren inversiones y recursos complementarios

Register for free at <https://www.scribd.com> to download the version without the watermark

4. Resultados

Se presentan los resultados de los análisis realizados. Siguiendo la métrica de Berners-Lee, la distribución de cada conjunto de datos se ha evaluado teniendo en cuenta los tres aspectos anteriores.

Como se observa en la tabla 2, la mayoría de las distribuciones (98,0%) tienen 2 estrellas (12,9%), que de forma habitual se corresponden con un fichero *Excel*, o 3 estrellas (85,2%) que en general se corresponden con fichero CSV. Es destacable que ninguna de las distribuciones tenga el nivel

Tabla 2. Características de la distribución de cada conjunto de datos

Estrellas	Número global de distribuciones	Número medio de descargas / mes
1	6	2,03
2	46	1,15
3	305	0,73
4	1	4,71
5	0	---

más alto (5 estrellas). Las 5 estrellas suponen que los datos están semantizados y además disponen de una ontología reconocida. Este resultado muestra que la semantización es una tarea pendiente y que para su implementación se requieren inversiones y recursos complementarios. Su aumento requiere un consenso del sector y/o del regulador en la elaboración de ontologías específicas.

Con relación a las descargas realizadas según el número de estrellas (tabla 2), se observa que en categorías menos frecuentes (1 y 4 estrellas), el número de descargas es mayor que en las categorías más frecuentes (2 y 3 estrellas). Entre las frecuentes se observa un descenso del 36,9% en número de descargas cuando se pasa de 2 a 3 estrellas. Este resultado no permite afirmar que una mayor calidad medida en términos del número de estrellas se corresponde con una mayor intensidad de utilización. No obstante podrían condicionar esta frecuencia de uso otros factores como la temática.

En el portal de datos abiertos de Barcelona sólo 12 de las 22 temáticas definidas en la NTI-RISP tienen representación entre los conjuntos de datos locales del portal. En la figura 2 se presentan los resultados obtenidos para el número medio de descargas por distribución y mes para cada temática en estas 12 categorías. Se observa que ciencia y tecnología es la temática con más descargas medias/mes, seguido por sociedad y bienestar, urbanismo e infraestructuras, cultura y ocio, y comercio. Se aprecia que la temática de los datos influye en su demanda. Por ejemplo, la demanda de la categoría de ciencia y tecnología es 26 veces mayor que la de demografía.

“Ciencia y tecnología es la temática con más descargas medias/mes, seguido por sociedad y bienestar, urbanismo e infraestructuras, cultura y ocio, y comercio”

Como se observa en la figura 3 hay un leve aumento en el número de descargas en función de la antigüedad de la última actualización del conjunto de datos. Este resultado abre la posibilidad de analizar con más profundidad esta relación ya que hay otros estudios que han encontrado que la actualización en tiempo real de los datos es un factor importante

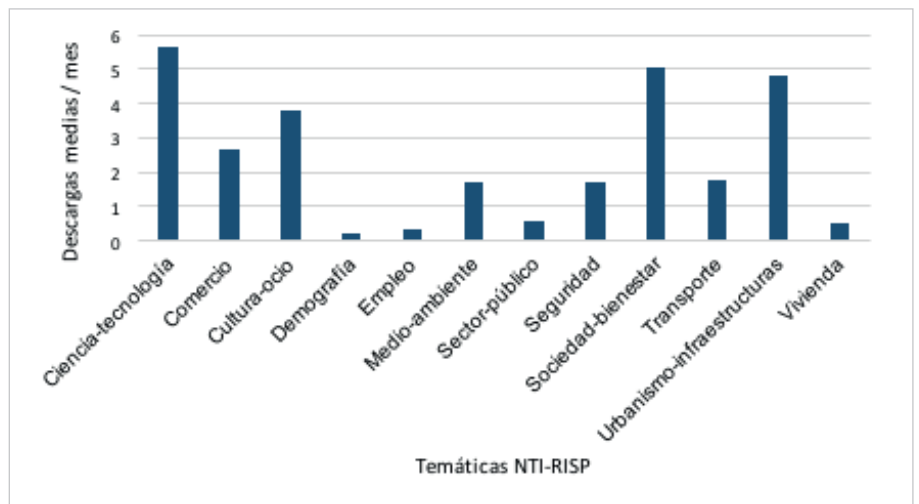


Figura 2. Temáticas y descargas

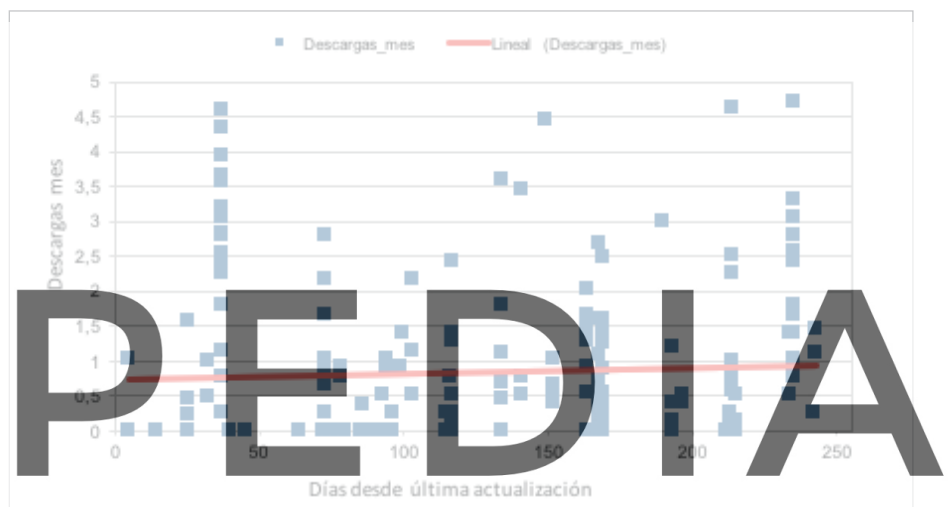


Figura 3. Descargas de distribuciones frente al tiempo desde la última actualización

Register for free at <https://www.scipedia.com> to download the version without the watermark

para el uso de servicios basados en los datos (Abella; Ortiz-de-Urbina-Criado; De Pablos-Heredero, 2017b). Otros trabajos sugieren que los datos abiertos incrementan su demanda con actualizaciones frecuentes (Kang; Son; Stankovic, 2004).

“El 10 de octubre de 2017 el portal de datos abiertos de Barcelona tenía 425 conjuntos de datos con 3.002 distribuciones”

En la figura 4 se observa que hay más descargas medias de los conjuntos de datos cuando la frecuencia de actualización es mayor. Se aprecia que la frecuencia de actualización es un factor que influye de forma directa en la demanda de los datos abiertos.

Como muestra la figura 5, la presencia de contenido geolocalizado multiplica por más de 20 el número de descargas. Sin embargo, cuando se tiene una previsualización de los datos (geolocalizados con mapa), las descargas sólo son unas 10 veces superiores a las de un conjunto de datos sin conteni-

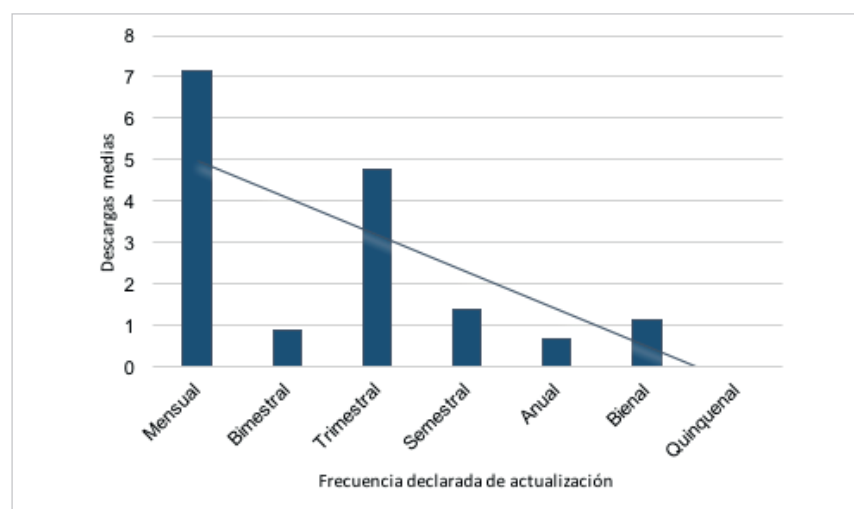


Figura 4. Descargas medias por frecuencia declarada de actualización



Figura 5. Descargas medias según contenido geográfico

Register for free at <https://www.scipedia.com> to download the version without the watermark

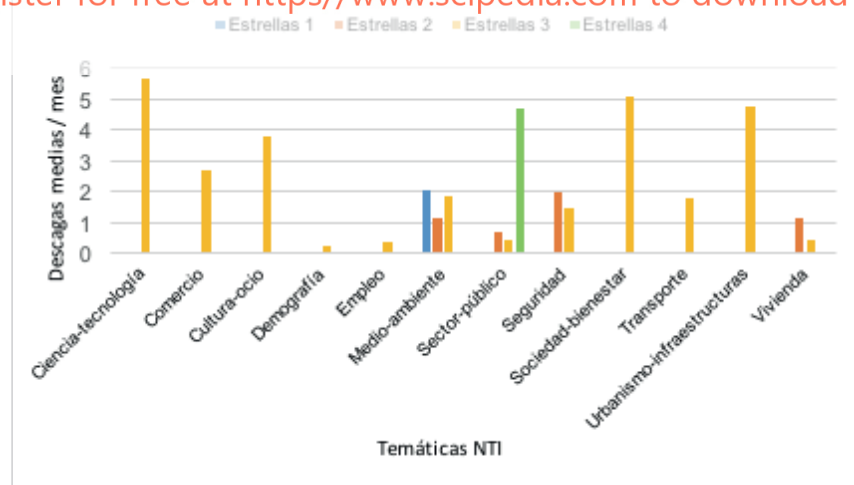


Figura 6. Descargas por temática y estrellas

do geográfico. Se aprecia que la geolocalización de los datos abiertos influye de forma directa en su demanda. Por otra parte este resultado indica que la visualización de los datos reduce la necesidad de descargarlos, lo que ofrece oportunidades de negocio (por ejemplo, el diseño de una aplicación para visualizar la ubicación de los autobuses de una ciudad).

formación en pocos conjuntos de datos más completos y manejables puede ayudar a mejorar la eficiencia en la publicación de los datos.

También se ha observado que hay diferencias entre las descargas directas y vía API; sin embargo en este trabajo no se disponía de datos sobre acceso vía API, siendo esta otra de

En la figura 6 se puede visualizar para los conjuntos de datos locales la evolución del número medio de descargas en función del número de estrellas de Berners-Lee para cada temática. En esta figura no se observa una relación entre el aumento de las estrellas y el número medio de descargas en categorías como medio ambiente, seguridad y vivienda. Predomina el formato de tres estrellas (normalmente, ficheros CSV); mientras que hay menos casos que tengan una o cuatro estrellas y ninguno con 5. Este resultado muestra que es necesario evolucionar los conjuntos de datos a niveles de 4 y 5 estrellas, ofreciendo así nuevas oportunidades para la creación de servicios.

5. Conclusiones

Autores como Janowicz *et al.* (2014) señalan que el modelo de Berners-Lee constituye una referencia para orientar a organizaciones e instituciones públicas en las posibilidades de publicar sus datos abiertos atendiendo a su acceso y potencial de uso. La investigación realizada muestra que si bien este modelo ofrece información sobre la apertura de los datos abiertos, hay otros factores que no incorpora y que pueden ayudar a evaluar mejor la calidad de los datos abiertos y su potencial para la reutilización.

Además, para evaluar los datos abiertos hay que tener en cuenta que la mera descarga no supone el uso de la información y además, la disponibilidad de herramientas de visualización de los datos hace que el usuario no necesite realizar tantas descargas, ya que puede consultar los datos online. Los trabajos de Elberse (2010) y Aguiar y Martens (2016) evidencian esta realidad centrándose en el efecto de la música online sobre las acciones de descarga de música.

Futuros estudios pueden analizar la relación entre la demanda de los datos abiertos y la dispersión de distribuciones y conjuntos de datos disponibles sobre una temática. Por ello sería interesante estudiar si concentrar la in-

las líneas de investigación futuras. Desde la perspectiva de los usuarios puede ser interesante contar con una interfaz de programación –API– que permita desarrollos profesionales incorporando los datos publicados. La herramienta tendría que proporcionar a los gestores del portal de datos abiertos información sobre lo que está pasando con el acceso a sus datos, tanto si el mecanismo de acceso es vía API o por descarga directa. Además sería conveniente crear algunos canales específicos para conocer las opiniones de los reutilizadores y los comentarios sobre el uso de los datos.

“La visualización de los datos reduce la necesidad de descargarlos, lo que ofrece oportunidades de negocio”

Muirragui-Irrazábal *et al.* (2016) definían algunas estadísticas de gestión de datos como el número de conjuntos de datos actualizados, la distribución de los conjuntos publicados, la frecuencia media de actualización y los formatos utilizados. En este trabajo se observa que la frecuencia de actualización se corresponde con un mayor número de descargas medias. Por ello se recomienda continuar analizando esta relación, así como otras estadísticas de gestión de datos.

Por otra parte se ha observado que la geolocalización es un factor que se asocia con un mayor número de descargas. Por tanto, resulta interesante realizar nuevos análisis de la visualización de datos para comprender la totalidad del uso. Otros trabajos como el de Čerba y Jedlička (2016) han llegado a conclusiones similares.

El alcance de esta investigación está limitado por el hecho de que no se han analizado los accesos vía API y que los datos estudiados provienen de una única fuente. Sin embargo supone un punto de partida para el desarrollo y mejora no sólo de la métrica de Berners-Lee sino también de otras métricas como *Meloda*. Asimismo los gestores de los portales de datos abiertos pueden extraer conclusiones útiles para mejorar sus procesos de publicación de datos, lo que puede facilitar el aumento de su demanda. Además la caracterización de los portales de datos puede ayudar a los infomediarios a identificar las mejores fuentes de datos para llevar a cabo sus servicios.

Agradecimientos

Agradecemos la colaboración y la información facilitada por el equipo del portal de datos abiertos de Barcelona.

Este trabajo ha sido financiado por el proyecto ECO2015-67434-R del *Ministerio de Economía y Competitividad* (España) y por el proyecto CEI, V638, del *Campus de Excelencia Internacional en Energía* de la *Universidad Rey Juan Carlos*.

6. Referencias

- Abella, Alberto; Ortiz-de-Urbina-Criado, Marta; De-Pablos-Heredero, Carmen (2014). “Meloda, métrica para evaluar la reutilización de datos abiertos”. *El profesional de la información*, v. 23, n. 6, pp. 582-588.
<https://doi.org/10.3145/epi.2014.nov.04>
- Abella, Alberto; Ortiz-de-Urbina-Criado, Marta; De-Pablos-Heredero, Carmen (2017a). “A model for the analysis of

data-driven innovation and value generation in smart cities’ ecosystems”. *Cities*, v. 64, pp. 47-53.

<https://goo.gl/fkc8rV>

<https://doi.org/10.1016/j.cities.2017.01.011>

Abella, Alberto; Ortiz-de-Urbina-Criado, Marta; De-Pablos-Heredero, Carmen (2017b). *La reutilización de datos abiertos: una oportunidad para España*. Fundación Cotec para la Innovación.

http://informecotec.es/media/INFORME_REUTILIZACION-DE-DATOS.pdf

Aguiar, Luis; Martens, Bertin (2016). “Digital music consumption on the internet: Evidence from clickstream data”. *Information economics and policy*, v. 34, pp. 27-43.

<https://doi.org/10.1016/j.infoecopol.2016.01.003>

Auer, Sören; Bizer, Christian; Kobilarov, Georgi; Lehmann, Jens; Cyganiak, Richard; Ives, Zachary (2007). “DBpedia: A nucleus for a web of open data”. *Lecture notes in computer science*, v. 4825, pp. 722-735.

https://doi.org/10.1007/978-3-540-76298-0_52

Berners-Lee, Tim (2006). *Linked data*. 27 July.

<https://www.w3.org/DesignIssues/LinkedData.html>

Bizer, Christian; Heath, Tom; Berners-Lee, Tim (2009). “Linked data. The story so far”. *International journal of semantic web and information systems*, v. 5, n. 3, pp. 1-22.

<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>

<https://doi.org/10.4018/jswis.2009081901>

Brin, Sergei; Page, Lawrence (1998). “The anatomy of a large-scale hypertextual web search engine”. *Computer networks*, v. 30, n. 1-7, pp. 107-117.

<http://infolab.stanford.edu/pub/papers/google.pdf>

[https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)

Čerba, Otakar; Jedlička, Karel (2016). “Linked forests: Semantic similarity of geographical concepts ‘forest’”. *Open geosciences*, v. 8, n. 1, pp. 556-566.

<https://doi.org/10.1515/geo-2016-0049>

Elberse, Anita (2010). “Bye-bye bundles: The unbundling of music in digital channels”. *Journal of marketing*, v. 74, n. 3, pp. 107-123.

http://www.people.hbs.edu/aelberse/publications/Elberse_2010.pdf

España (2013). “Resolución de 19 de febrero de 2013, de la Secretaría de Estado de Administraciones Públicas, por la que se aprueba la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información”. *BOE*, n. 54, 4 marzo.

https://www.boe.es/diario_boe/txt.php?id=BOE-A-2013-2380

Ferrer-Sapena, Antonia; Peset, Fernanda; Aleixandre-Benavent, Rafael (2011). “Acceso a los datos públicos y su reutilización: open data y open government”. *El profesional de la información*, v. 20, n. 3, 260-269.

<https://doi.org/10.3145/epi.2011.may.03>

Garriga-Portolà, Marc (2011). “¿Datos abiertos? Sí, pero de forma sostenible”. *El profesional de la información*, v. 20, n. 3, pp. 298-303.

SCIPEDIA

Register for free at <https://www.scipedia.com> to download the version without the watermark

<https://doi.org/10.3145/epi.2011.may.08>

Gerhard, Tim; Schwerdel, Denis; Müller, Paul (2017). "Providing information centrality in today's internet via software-defined networking and in-network processing". *PIK-praxis der informations verarbeitung und kommunikation*, v. 39, n. 3-4, pp. 71-86.

<https://goo.gl/pyPFEx>

<https://doi.org/10.1515/pik-2017-0002>

Gértrudix, Manuel; Gertrudis-Casado, María-Carmen; Álvarez-García, Sergio (2016). "Consumption of public institutions' open data by Spanish citizens". *El profesional de la información*, v. 25, n. 4, pp. 535-544.

<https://doi.org/10.3145/epi.2016.jul.03>

Hogan, Mél; Shepherd, Tamara (2015). "Information ownership and materiality in an age of big data surveillance". *Journal of information policy*, v. 5, pp. 6-31.

<https://doi.org/10.5325/jinfopoli.5.2015.0006>

Janowicz, Krzysztof; Hitzler, Pascal; Adams, Benjamin; Kolas, Dave; Vardeman, Charles (2014). "Five stars of linked data vocabulary use". *Semantic web*, n. 0, pp. 1-3.

<http://www.semantic-web-journal.net/content/five-stars-linked-data-vocabulary-use>

Janssen, Marijn; Charalabidis, Yannis; Zuiderwijk, Anneke (2012). "Benefits, adoption barriers and myths of open data and open government". *Information systems management*, v. 29, n. 4, pp. 258-268.

<https://goo.gl/d3NJUE>

<https://doi.org/10.1080/10580530.2012.716740>

Kalampokis, Evangelos; Tambouris, Efthimios; Tarabanis, Konstantinos (2011). "A classification scheme for open government data: towards linking decentralised data". *International journal of web engineering and technology*, v. 6, n. 3, pp. 266-285.

<https://goo.gl/oxi4Lu>

<https://doi.org/10.1504/IJWET.2011.040725>

Kang, Kyoung-Don; Son, Sang H.; Stankovic, John A. (2004). "Managing deadline miss ratio and sensor data freshness in real-time databases". *IEEE transactions on knowledge and data engineering*, v. 16, n. 10, pp. 1200-1216.

<https://doi.org/10.1109/TKDE.2004.61>

Marcos-Martín, Carlos; Soriano-Maldonado, Salvador-Luis (2011). "Reutilización de la información del sector público y open data en el contexto español y europeo. Proyecto Aporta". *El profesional de la información*, v. 20, n. 3, pp. 291-297.

<http://dx.doi.org/10.3145/epi.2011.may.07>

Martínez-Méndez, Francisco-Javier; López-Carreño, Rosana; Pastor-Sánchez, Juan-Antonio (2014). "Open data en las administraciones públicas españolas: categorías temáticas y apps". *El profesional de la información*, v. 23, n. 4, pp. 415-423.

<https://doi.org/10.3145/epi.2014.jul.10>

Maseda-Seco, Diego; Bueno-de-la-Fuente, Gema; Méndez, Eva (2017). "Análisis y categorización de los datos abiertos de las bibliotecas municipales españolas: metadatos, interoperabilidad y propuesta para la apertura y reutilización". *El profesional de la información*, v. 26, n. 3, pp. 392-402.

<https://doi.org/10.3145/epi.2017.may.05>

Muirragui-Irrazábal, Viena; Pacheco-Olea, Fernando; León-Plúas, Edwin; Pacheco-Olea, Leonidas (2016). "Las plataformas de publicación: herramientas open data". *Revista ciencias pedagógicas e innovación*, v. 4, n. 1, pp. 61-68.

<https://doi.org/10.26423/rcpi.v4i1.117>

Oviedo, Edgar; Mazón, Jose-Norberto; Zubcoff, Jose-Jacobo (2013). "Towards a data quality model for open data portals". En: *XXXIX Latin American computing conference (CLEI)*. <https://dx.doi.org/10.1109/CLEI.2013.6670665>.

Piedra, Nelson; Chicaiza, Janneth; López, Jorge; Tovar-Caro, Edmundo (2017). "A rating system that open-data repositories must satisfy to be considered OER: Reusing open data resources in teaching". En: *2017 IEEE Global engineering education conf (Educon)*, pp. 1768-1777.

<https://goo.gl/YZmrd4>

<https://doi.org/10.1109/EDUCON.2017.7943089>

Ribeiro, Cristiano E.; Vivacqua, Adriana S. (2013). "A framework for composition and reuse on the linked open data cloud". En: *IEEE 7th Intl conf on semantic computing*.

<http://dx.doi.org/10.1109/ICSC.2013.14>

Shankar, Kalpana (2015). "For want of a nail: Three tropes in data curation". *Preservation, digital technology & culture*, v. 44, n. 4, pp. 161-170.

<http://researchrepository.ucd.ie/handle/10197/7272>

<https://doi.org/10.1515/pdtc-2015-0019>

Sowe, Sulayman K.; Zettsu, Koji (2015). "Towards an open data development model for linking heterogeneous data sources". *7th Intl conf on knowledge and systems engineering*, pp. 344-347.

<https://doi.org/10.1109/KSE.2015.56>

Tambassi, Timothy (2016). "Rethinking geo-ontologies from a philosophical point of view". *Journal of research and dialectics in geography (J-Reading)*, v. 2, n. 3, pp. 51-62.

<https://goo.gl/fkjiHg>

Vicente-Paños, Adrián; Jordán-Alfonso, Aurea (2017). "Acceso a la información pública y su reutilización en las comunidades autónomas: evaluación de la reutilización de datos abiertos". *El profesional de la información*, v. 26, n. 3, pp. 381-391.

<https://doi.org/10.3145/epi.2017.may.04>

Weiskopf, Nicole G.; Weng, Chunhua (2013). "Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research". *Journal of the American Medical Informatics Association*, v. 20, n. 1, pp. 144-151.

<https://doi.org/10.1136/amiajnl-2011-000681>

Yang, Hsin-Chang; Lin, Cathy S.; Yu, Po-Han (2015). "Toward automatic assessment of the categorization structure of open data portals". En: Wang, Leon; Uesugi, i-Hsien-Ting; Okuhara, Koji; Wang, Kai. *Multidisciplinary social networks research*. MISNC 2015. Berlin, Heidelberg: Springer. ISBN: 978 3 662 48318 3

https://doi.org/10.1007/978-3-662-48319-0_30

Zuiderwijk, Anneke; Jeffery, Keith; Janssen, Marijn (2012). "The potential of metadata for linked open data and its value for users and publishers". *Journal of edemocracy & open government*, v. 4, n. 2, pp. 222-244.

<http://www.jedem.org/index.php/jedem/article/view/138>

SCIPEDIA

Register for free at <https://www.scipedia.com> to download the version without the watermark